ΕΙΣΑΓΩΓΗ ΣΤΗΝ ΣΤΑΤΙΣΤΙΚΗ

29/03/2018

Διάλεξη 11η (Σημαντικό μάθημα για το τέλος)

Συμπερασματολογία με την διαφορά δύο μέσων τιμών ($\mu_1-\mu_2$)

Παράδ. 4.5

Δύο ανεξάρτητες ομάδες παιδιών αποτελούνται από 10 παιδιά η κάθε μία. Τα παιδιά των υπερτασικών γονέων ανήκουν στην ομάδα 1, ενώ τα παιδιά των κανονικών γονέων ανήκουν στην ομάδα 2.
Η συστολική πίεση του αίματος μετριέται στα παιδιά των 2 ομάδων με τα ακόλουθα αποτελέσματα:

($n_1=10$) ομάδα 1: 90, 112, 100, 102, 96, 106, 110, 110, 120, 112

($n_2=10$) ομάδα 2: 104, 88, 100, 98, 102, 96, 100, 96, 96, 92

Έστω: $\underbrace{N(\mu_1,\sigma_1^2)\ ,\ N(\mu_2,\sigma_2^2)}_{\text{κανονική κατανομή}}$ και $\sigma_1^2=\sigma_2^2$

$\mu_1=\mu_2$;

Δείγμα	Μέγεθος	Βαθμός ελευθερίας (ΒΕ)	Μέση τιμή	Αθροίσματα τετραγώνων
1	$n_1=10$	$n_1-1=\boxed{9}$	$\bar{x}_1=105.8$	$\sum_i^{n_1}(x_{1i}-\bar{x}_1)^2=\boxed{707.6}$
2	$n_2=10$	$n_2-1=\boxed{9}$	$\bar{x}_2=97.2$	$\sum_i^{n_2}(x_{2i}-\bar{x}_2)^2=\boxed{201.6}$
		$n_1+n_2-2=18$		

Συνολικό S.S $=909.2$

$H_0: \mu_1-\mu_2=0$

$H_a: \mu_1-\mu_2\neq 0$

$$t=\frac{\bar{x}_1-\bar{x}_2-0}{S_p\sqrt{\frac{1}{10}+\frac{1}{10}}}=2.70 \qquad S_p=\sqrt{\frac{909.2}{18}}=7.11$$

κρίσιμη περιοχή $|t|\geq t_{0{,}025,18}=2.101$ απορρ. η H_0.

Γενικά: Έστω $X_{11}, \dots, X_{1n_1}$ τ.δ. από πληθ. 1 (μ_1, σ_1^2)
και $X_{21}, \dots, X_{2n_2}$ τ.δ. από άλλο πληθ. 2 (μ_2, σ_2^2)

Μας ενδιαφέρει η παράμετρος $\mu_1 - \mu_2$.

$$\boxed{\bar{X}_1 = \frac{1}{n_1}\sum_{i=1}^{n_1} X_{1i}}, \quad \boxed{\bar{X}_2 = \frac{1}{n_2}\sum_{i=1}^{n_2} X_{2i}},$$

$$\boxed{S_1^2 = \frac{1}{n_1-1}\sum_{i=1}^{n_1}(X_{1i}-\bar{X}_1)^2}, \quad \boxed{S_2^2 = \frac{1}{n_2-1}\sum_{i=1}^{n_2}(X_{2i}-\bar{X}_2)^2}$$

Δείγμα	Μέγεθος	Βαθμός ελευθερίας (Β.Ε)	Μέση Τιμή	Αθροίσματα Τετραγώνων
1	n_1	n_1-1	$\bar{X}_1$	$\sum_i^{n_1}(X_{1i}-\bar{X}_1)^2$
2	n_2	n_2-1	$\bar{X}_2$	$\sum_i^{n_2}(X_{2i}-\bar{X}_2)^2$

Για την παράμετρο $\mu_1-\mu_2$:

$$\widehat{\mu_1-\mu_2} = \bar{X}_1 - \bar{X}_2,$$

$$\sigma^2_{\bar{X}_1-\bar{X}_2} = Var(\bar{X}_1-\bar{X}_2) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2} \quad \text{και}$$

$$\hat{\sigma}^2_{\bar{X}_1-\bar{X}_2} = \frac{S_1^2}{n_1} + \frac{S_2^2}{n_2} \quad \text{και} \quad \hat{\sigma}_{\bar{X}_1-\bar{X}_2} = \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}.$$

(Ι) Κανονικοί πληθυσμοί, "γνωστά" σ_1^2, σ_2^2

$$\left.\begin{aligned} \bar{X}_1 &\sim N\left(\mu_1, \frac{\sigma_1^2}{n_1}\right) \\ \bar{X}_2 &\sim N\left(\mu_2, \frac{\sigma_2^2}{n_2}\right) \end{aligned}\right\} \Rightarrow \bar{X}_1 - \bar{X}_2 \sim N\left(\mu_1-\mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)$$

$$\frac{\bar{x}_1 - \bar{x}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,1) \Rightarrow$$

$(1-a)100\%$ Δ.Ε. για το $\mu_1 - \mu_2$: $\bar{x}_1 - \bar{x}_2 \pm z_{\frac{a}{2}} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$

$\rightsquigarrow$ Για να ελέγξουμε υποθέσεις της μορφής:

(i) $H_0 : \mu_1 - \mu_2 \leq \delta_0$ (γνωστό) v $H_a : \mu_1 - \mu_2 > \delta_0$: $C = [z_a, \infty)$ ή $z \geq z_a$

(ii) $H_0 : \mu_1 - \mu_2 \geq \delta_0$ (γνωστό) v $H_a : \mu_1 - \mu_2 < \delta_0$: $C = (-\infty, -z_a]$ ή $z \leq -z_a$

(iii) $H_0 : \mu_1 - \mu_2 = \delta_0$ v $H_a : \mu_1 - \mu_2 \neq \delta_0$: $C = (-\infty, -z_{\frac{a}{2}}] \cup [z_{\frac{a}{2}}, \infty)$ ή $|z| \geq z_{\frac{a}{2}}$

Χρησιμοποιείται το στατιστικό:

$$\boxed{z = \frac{\bar{x}_1 - \bar{x}_2 - \delta_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}}$$

και για επίπεδο σημαντικότητας a :

$C = [z_a, \infty)$ ή $z \geq z_a$

$C = (-\infty, -z_a]$ ή $z \leq -z_a$

$C = (-\infty, -\frac{z_a}{2}] \cup [\frac{z_a}{2}, \infty)$ ή $|z| \geq \frac{z_a}{2}$

(II) Κανονικοί Πληθυσμοί : $\sigma_1^2 = \sigma_2^2 = \sigma^2$ άγνωστο

$$\left.\begin{array}{l}\bar{X}_1 \sim N\left(\mu_1, \frac{\sigma^2}{n_1}\right) \\ \bar{X}_2 \sim N\left(\mu_2, \frac{\sigma^2}{n_2}\right)\end{array}\right\} \Rightarrow \bar{X}_1 - \bar{X}_2 \sim N\left(\mu_1 - \mu_2, \sigma^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\right)$$

$$\frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t_{n_1+n_2-2}$$

$$S_p^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2} = \frac{\text{Συνολικό S.S.}}{\text{Συνολικοί B.E.}}$$

Έτσι για ένα $(1-\alpha)100\%$ Δ.Ε για την $\mu_1 - \mu_2$:

$$\boxed{\bar{X}_1 - \bar{X}_2 \pm t_{\frac{\alpha}{2}, n_1+n_2-2} \cdot S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

Χρησιμοποιούμε το στατιστικό $\boxed{t = \frac{\bar{X}_1 - \bar{X}_2 - \delta_0}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \overset{H_0}{\sim} t_{n_1+n_2-2}}$

και για επίπεδο σημαντικότητας α:

$t \geq t_{\alpha, n_1+n_2-2}$ (για την (i))

$t \leq -t_{\alpha, n_1+n_2-2}$ (για την (ii))

$|t| \geq t_{\alpha/2, n_1+n_2-2}$ (για την (iii))

Παράδ. 4.6 (Επεξ. θερμότητας)

Είναι η επεξεργ. θερμότητας αποτελεσματική; ($\alpha=0{,}05$)

Δείγμα:	1	2	3	4	5	6	7	8	9	10
Πριν (x_{1i})	6.88	7.05	8.94	5.90	6.16	6.67	7.01	5.46	5.87	6.54
Μετά (x_{2i})	6.85	6.84	7.07	5.05	6.18	6.71	6.49	5.94	5.88	6.41
Διαφορές $D_i = x_{1i}-x_{2i}$	0,03	0.21	1.17	0.85	-0.02	-0,04	0,52	0.52	-0,01	0.13

$H_0: \delta=\delta_0 \quad v \quad H_a: \delta>\delta_0 \ (\delta_0=0)$ | $H_0: \delta=0 \ v \ H_a: \delta>0$ για επίλυση.

$D_i \sim N(\delta, \sigma^2$ άγνωστο$), \ i=1,\dots,n$

$$t=\frac{\bar{D}-\delta_0}{s_D/\sqrt{n}} \overset{H_0}{\sim} t_{n-1}.$$

$$t=\frac{\bar{D}-0}{s_D/\sqrt{n}}=\frac{0.236}{0.366/\sqrt{10}}=2.03$$

Επειδή $2.03>1.833$ απορρ. H_0.

Δηλαδή η επεξεργασία λοιπόν είναι αποτελεσματική.

κρισ. περ. $t \geq t_{0,05,9}=(1.833)$